A class of augmented Lagrangian algorithms for infinite-dimensional optimization with equality constraints

Sachs, E.; Sartenaer, Annick

Publication date:
2005

Document Version
Peer reviewed version

Link to publication

Citation for pulished version (HARVARD):
Sachs, E & Sartenaer, A 2005, *A class of augmented Lagrangian algorithms for infinite-dimensional optimization with equality constraints*. FUNDP, Faculté des Sciences. Département de Mathématique., Namur.

A class of augmented Lagrangian algorithms for infinite-dimensional optimization with equality constraints

E.W. Sachs* Annick Sartenaer†

July 13, 2005

Abstract

We consider a class of augmented Lagrangian algorithms for the solution of optimization problems posed in an *infinite-dimensional* setting. This class extends the augmented Lagrangian algorithm developed in [2] (when considering equality constraints only) which is motivated by solving a sequence of subproblems in which the augmented Lagrangian is *approximately* minimized – i.e., each subproblem is terminated as soon as a stopping condition is satisfied. Global and local convergence results are outlined. A study of the behavior of the extended class of algorithms is further presented when the sequence of approximately solved infinite-dimensional subproblems is replaced by a sequence of finite-dimensional subproblems obtained by a more and more refined discretization of their infinite-dimensional counterpart.

Key words: Nonlinear optimization, equality constraints, infinite-dimensional optimization, augmented Lagrangian methods, discrete approximations

1 Introduction

The problem we consider in this paper is that of calculating a local minimizer of a smooth function subject to general equality constraints. That is, we wish to solve the problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad c(x) = 0, \quad x \in X, \tag{1.1}$$

where f and c are maps defined as follows

$$f : X \to I\!R, \qquad c : X \to Y$$

with Hilbert spaces X and Y over the reals. A classical technique for solving problem (1.1) is to minimize a suitable sequence of *augmented Lagrangian functions.* These functions are

* FB IV - Mathematik, University of Trier, D-54286 Trier, Germany, email: sachs@uni-trier.de and Department of Mathematics, Virginia Tech, email: sachs@icam.vt.edu.

† Department of Mathematics, Facultés Universitaires Notre-Dame de la Paix, 61, rue de Bruxelles, B-5000 Namur, Belgium, email: annick.sartenaer@fundp.ac.be.

defined by $\Phi : X \times Y \times I\!R \to I\!R$

$$\Phi(x, \lambda, \mu) = f(x) + < \lambda, c(x) >_Y + \frac{1}{2\mu}\|c(x)\|_Y^2, \tag{1.2}$$

where the vector λ is known as a *Lagrange multiplier estimate* and μ is known as the *penalty parameter* (see, for instance, Bertsekas [1]). $< \cdot, \cdot >_Z$ denotes the inner product in some Hilbert space Z, and let $\|\cdot\|_Z =< \cdot, \cdot >_Z^{1/2}$ (or simply $\|\cdot\|$ when it is clear from the context) be the corresponding norm.

In this paper, we base the convergence results on [2] applied to the Hilbert space case. In order to apply this to a sequence of discretized problems, we introduce a new set-valued update rule for Lagrange multipliers and an algorithm based refinement rule for the discretization. In contrast to [2] we omit linear inequality constraints and use real valued penalty parameters.

Other papers that deal with augmented Lagrangian methods in Hilbert space were written by Ito and Kunisch (see [4] and [5]), where local and global convergence results were given, however for fixed penalty parameters. Volkwein [7] presents in his dissertation a Kantorovich theory which allows to derive mesh independence results for augmented Lagrangian methods. In a paper by Sachs and Volkwein [6] these results were extended to more general update procedures for the Lagrange multiplier.

When solving an infinite-dimensional optimization problem one often uses a sequence of finite-dimensional discretized problems. Rather than solving these accurately at each level of discretization, we want to start with a coarse level and increase this level as the iteration progresses. Then the question arises how this is coupled with the penalty parameter and the Lagrange multiplier update. We interpret each iterate as an iteration in an algorithm in infinite dimensions. This can be done because each subproblem in the infinite-dimensional counterpart needs not to be solved exactly allowing the discretization error being interpreted as inexactness. This basic principle required us to introduce also a new update rule for penalty parameter and Lagrange multiplier. The level of refinement is determined in the algorithm based on the progress towards stationarity.

The problem and the notation are introduced in §2. Section 3 presents the proposed algorithm including set-valued update rules and inexact solves of the minimization of the augmented Lagrangian. The global and local convergence analysis in Hilbert space is developed in §4. In §5.1, we consider a sequence of discretized problems and define an algorithm where at each step the discretization is refined if necessary. The resulting sequence of iterates is then interpreted as an iteration in an infinite-dimensional setting so that we can apply the convergence theorems from §4. In §5.2, we use the framework of restrictions and projections to define approximating functions and check the previously stated assumptions with these discretized optimization problems.

2 The problem and notation

We consider the problem stated in (1.1). On the functions we make the following assumptions.

AS1: The functions f and c are twice continuously Fréchet-differentiable.

We now introduce the notation that will be used throughout the paper. For any function F let $\nabla_x F(x)$ denote the gradient of $F(x)$ and $F'(x), F''(x)$ denote the operator defined by the first and second derivative of F, respectively. The Lagrangian function is given by $\Phi(x, \lambda, 0)$ or

$$\ell(x, \lambda) = f(x) + < \lambda, c(x) >_Y .$$

If we define the *first-order Lagrange multiplier estimate* as

$$\bar{\lambda}(x, \lambda, \mu) = \lambda + c(x)/\mu, \tag{2.1}$$

we shall use the identity

$$\begin{aligned} \nabla_x \Phi(x, \lambda, \mu) &= \nabla_x f(x) + c'(x)^* \lambda + \frac{1}{\mu} c'(x)^* c(x) \\ &= \nabla_x f(x) + c'(x)^* \bar{\lambda}(x, \lambda, \mu) \\ &= \nabla_x \ell(x, \bar{\lambda}(x, \lambda, \mu)). \end{aligned} \tag{2.2}$$

If the map

$$c'(x) : X \rightarrow Y \quad \text{is surjective},$$

then the map $c'(x)c'(x)^* : Y \rightarrow Y$ is surjective and injective and therefore has a continuous inverse. Hence its generalized right inverse exists and we define the *least-squares Lagrange multiplier estimate*

$$\lambda(x) \stackrel{\text{def}}{=} -(c'(x)^+)^* \nabla_x f(x) \tag{2.3}$$

where the generalized right inverse is given by

$$c'(x)^+ \stackrel{\text{def}}{=} c'(x)^* (c'(x) c'(x)^*)^{-1}.$$

We now describe more precisely the algorithm that we propose to use.

3 Description of the algorithm

As in [2], the algorithmic model we propose to solve problem (1.1) proceeds at iteration k by computing an iterate x_k that approximately solves the subproblem

$$\min_{x \in X} \Phi(x, \lambda_k, \mu_k), \tag{3.1}$$

where the values of the Lagrange multiplier estimate λ_k and the penalty parameter μ_k are fixed for the subproblem. Subsequently we update the Lagrange multiplier or decrease the penalty parameter, depending on how much the constraint violation has been reduced. The tests on the size of the constraint violation are designed to allow the multiplier updates to take over in the neighbourhood of a stationary point.

The approximate minimization for problem (3.1) is performed in an *inner iteration* which is stopped as soon as its current iterate is sufficiently critical, in the sense that

$$\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\| \leq \omega_k,$$

where ω_k is a suitable tolerance at iteration k decreasing to zero with $k \to \infty$.

Algorithm ALINF below follows closely the algorithmic model described in [2], but introduces two relaxations, one in the tests on the size of the constraint, and another in the update of the Lagrange multiplier estimate. These two relaxations will allow the solution of problem (1.1) to be computed by considering a sequence of *finite-dimensional subproblems*, obtained by a more and more refined discretization of their infinite-dimensional counterpart. Indeed, we will first show in §4 that the proposed relaxations do not spoil the convergence results derived for the algorithmic model of [2]. We will then interpret, thanks to the relaxations, the iterates generated by the sequence of finite-dimensional subproblems as an iteration in the infinite-dimensional setting given by our algorithmic model (see §5.1). As a consequence of this, the convergence theory developed in §4 will then apply to the sequence of solutions computed from the discretized subproblems.

To motivate this paper in more detail, consider the algorithm of [2]: The decision rule in Step 3 of that algorithm which is used to decide whether either (case a) the Lagrange multiplier is updated or (case b) the penalty parameter is reduced has the following form:

$$\begin{array}{llll} \text{if} & \|c(x_k)\| & < & \eta_k \quad \text{goto case a,} \\ \text{if} & \|c(x_k)\| & \geq & \eta_k \quad \text{goto case b,} \end{array}$$

where η_k is decreasing to zero. If one analyzes a sequence of approximating problems where the constraint c is approximated by c_n, $n \in I\!N$, say, then it is conceivable that for a fixed iterate x_k the sequence of function values $c_n(x_k)$ converges to $c(x_k)$ under appropriate assumptions. However, it could happen that the values $\|c_n(x_k)\|$ oscillate around η_k and the decision for case a or case b can alternate as n tends to infinity. Note that this could happen if $\|c(x_k)\| = \eta_k$. This could lead to different sequences of iterates even when c_n approaches c. Therefore, this leads us to revise the decision rule in Step 3 of the algorithm in such a way that we introduce a fuzzy area for the decision: For $0 < \gamma_1 < 1 < \gamma_2$ we require that

$$\begin{array}{lllll} \text{if} & \|c(x_k)\| & \leq & \gamma_1\eta_k & \text{goto case a,} \\ \text{if} & \|c(x_k)\| & \in & (\gamma_1\eta_k, \gamma_2\eta_k) & \text{goto case a or b,} \\ \text{if} & \|c(x_k)\| & \geq & \gamma_2\eta_k & \text{goto case b,} \end{array}$$

is executed.

The second change in the algorithm is an appoximation of the Lagrange multiplier update. The requirement is that it has to be close to the first order multiplier update up to an error ω_k:

$$\|\lambda_{k+1} - \bar{\lambda}(x_k, \lambda_k, \mu_k)\| \leq \omega_k.$$

We are now ready to define our algorithmic model in more detail.

Algorithm ALINF

Step 0 [Initialization]. An initial Lagrange multiplier estimate $\lambda_0 \in Y$ and a positive penalty parameter $\mu_0 < 1$ are given. The positive constants $\omega_* < 1$, $\eta_* < 1$, $\gamma_1 < 1$, $\gamma_2 > 1$, $\tau < 1$, $\alpha_\eta < 1$, and $\beta_\eta < 1$ are specified. Set $\omega_0 = \mu_0$, $\eta_0 = \mu_0^{\alpha_\eta}$, and $k = 0$.

Step 1 [Inner iteration]. Find $x_k \in X$ that approximately solves (3.1), i.e., such that

$$\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\| \leq \omega_k \tag{3.2}$$

holds.

Step 2 [Test for convergence]. If $\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\| \leq \omega_*$ and $\|c(x_k)\| \leq \eta_*$, stop.

Step 3 [Updates]. If

$$\|c(x_k)\| \leq \gamma_1 \eta_k, \tag{3.3}$$

execute Step 3a. If

$$\|c(x_k)\| \geq \gamma_2 \eta_k, \tag{3.4}$$

execute Step 3b. Otherwise if

$$\gamma_1 \eta_k < \|c(x_k)\| < \gamma_2 \eta_k, \tag{3.5}$$

execute Step 3a or Step 3b. Increment k by one and go to Step 1.

Step 3a [Update Lagrange multiplier estimate]. Choose λ_{k+1} that satisfies

$$\|\lambda_{k+1} - \bar{\lambda}(x_k, \lambda_k, \mu_k)\| \leq \omega_k \tag{3.6}$$

and set

$$\begin{aligned} \mu_{k+1} &= \mu_k, \\ \omega_{k+1} &= \omega_k \mu_{k+1}, \\ \eta_{k+1} &= \eta_k \mu_{k+1}^{\beta_\eta}. \end{aligned} \tag{3.7}$$

Step 3b [Reduce the penalty parameter]. Set

$$\begin{aligned} \lambda_{k+1} &= \lambda_k, \\ \mu_{k+1} &= \tau \mu_k, \\ \omega_{k+1} &= \mu_{k+1}, \\ \eta_{k+1} &= \mu_{k+1}^{\alpha_\eta}. \end{aligned} \tag{3.8}$$

Note that, as in [2], Algorithm ALINF is coherent, in that

$$\lim_{k \to \infty} \omega_k = \lim_{k \to \infty} \eta_k = 0. \tag{3.9}$$

We will see in §5 how this algorithm can be used to prove convergence for a method which is adaptively changing the discretization levels as the iteration progresses.

4 Convergence analysis for Algorithm ALINF

The convergence theory is based on that in [2] and [3]. In contrast to the problem formulation in these papers, we do not include bound constraints [3] or linear inequality constraints [2]. However, we formulate the problem in a infinite-dimensional setting. On the other hand, the algorithm considered in this paper includes a *more general decision rule* for the updates of the Lagrange multiplier, the penalty parameter and the tolerances, as well as an *inexact updating rule* for the Lagrange multiplier. We moved some of the proofs of the convergence statements into the appendix.

4.1 Global Convergence analysis

We will show that Algorithm ALINF is globally convergent under the following assumptions.

AS2: The iterates $\{x_k\}$ lie within a compact set Ω.

AS3: The operator $c'(x_*)$ is surjective at any limit point, x_*, of the sequence $\{x_k\}$.

Assumption AS2 implies that there exists at least a convergent subsequence of iterates, while Assumption AS3 guarantees that the Fréchet-derivative of the constraint is surjective in an neighborhood of x^*, and we suppose w.l.o.g. that $c'(x_k)$ is surjective for the sequence of iterates.

For the purpose of our convergence analysis, we assume that the convergence tolerances ω_* and η_* are both zero. We now state the analog of Lemma 4.4 in [2].

Lemma 4.1 *Suppose that AS1 holds. Let $\{x_k\}, k \in \mathcal{K}$, be a sequence satisfying AS2 which converges to the point x_* for which AS3 holds and let $\lambda_* = \lambda(x_*)$, where λ satisfies (2.3). Assume that $\{\lambda_k\}$, $k \in \mathcal{K}$, is any sequence of vectors and that $\{\mu_k\}$, $k \in \mathcal{K}$, form a non-increasing sequence of scalars. Suppose further that (3.2) holds where the ω_k are positive scalar parameters which converge to zero as $k \in \mathcal{K}$ increases. Then*

(i) There are positive constants κ_1 and κ_2 such that

$$\|\bar{\lambda}(x_k, \lambda_k, \mu_k) - \lambda_*\| \le \kappa_1 \omega_k + \kappa_2 \|x_k - x_*\|, \tag{4.1}$$

$$\|\lambda(x_k) - \lambda_*\| \le \kappa_2 \|x_k - x_*\|, \tag{4.2}$$

and

$$\|c(x_k)\| \le \kappa_1 \omega_k \mu_k + \mu_k \|(\lambda_k - \lambda_*)\| + \kappa_2 \mu_k \|x_k - x_*\|, \tag{4.3}$$

for all $k \in \mathcal{K}$ sufficiently large.

Suppose, in addition, that $c(x_*) = 0$. Then

(ii) x_* is a Karush-Kuhn-Tucker point (first-order stationary point) for the problem (1.1), λ_* is the corresponding Lagrange multiplier, and the sequences $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ and $\{\lambda(x_k)\}$ converge to λ_* for $k \in \mathcal{K}$;

(iii) The gradients $\nabla_x \Phi(x_k, \lambda_k, \mu_k)$ converge to $\nabla_x \ell(x_*, \lambda_*)$ for $k \in \mathcal{K}$.

The detailed proof of Lemma 4.1 is given in the appendix, and does not need to take the two relaxations introduced in Algorithm ALINF into consideration. This proof can be easily derived from the proof of Lemma 4.4 in [2], using two steps. The first step consists in simplifying the proof of Lemma 4.4 in [2] by considering only one penalty parameter μ_k and by assuming $Z_* = I$ (where I is the identity operator) and A_* equals zero, since we do not consider linear inequality constraints in our framework. The second step takes account of the infinite-dimensional setting in which problem (1.1) is posed, and replaces J_k^T in [2] by

$c'(x_k)^*$. Note that inequality (4.2), which is the equivalent of inequality (4.14) in [2], holds because of the Lipschitz continuity of $c'(x)$ in a neighborhood of x_*, and hence of $c'(x)^+$.

We now require the following lemma in the proof of the global convergence, which shows that the Lagrange multiplier estimate cannot behave too badly. Again this lemma is adapted from the global convergence theory in [2] (see Lemma 4.5), but since the two relaxations intoduced in Algorithm ALINF interfere explicitly, we detail the proof.

Lemma 4.2 *Suppose that μ_k converges to zero as k increases when Algorithm ALINF is executed. Then the product $\mu_k\|\lambda_k\|$ converges to zero.*

Proof. As μ_k converges to zero, Step 3b must be executed infinitely often. Let $\mathcal{K} = \{k_0, k_1, k_2, ...\}$ be the set of indices of the iterations in which Step 3b is executed and for which

$$\mu_k \leq \left(\frac{1}{2}\right)^{1/\beta_\eta} \leq \frac{1}{2}. \tag{4.4}$$

We consider how the Lagrange multiplier estimate change between two successive iterations indexed in the set $\mathcal{K}$. At iteration $k_i + j$, for $k_i < k_i + j \leq k_{i+1}$, we have, since $\lambda_{k_i+1} = \lambda_{k_i}$ and by (2.1)

$$\begin{aligned} \lambda_{k_i+j} &= \lambda_{k_i+1} + \textstyle\sum_{l=1}^{j-1} \left[\lambda_{k_i+l+1} - \lambda_{k_i+l}\right] \\ &= \lambda_{k_i} + \textstyle\sum_{l=1}^{j-1} \left[\lambda_{k_i+l+1} - \bar{\lambda}(x_{k_i+l}, \lambda_{k_i+l}, \mu_{k_i+l}) + c(x_{k_i+l})/\mu_{k_i+l}\right] \\ &= \lambda_{k_i} + \textstyle\sum_{l=1}^{j-1} \left[c(x_{k_i+l})/\mu_{k_i+l} + e_{k_i+l}\right], \end{aligned} \tag{4.5}$$

where we use the notation $e_k \overset{\text{def}}{=} \lambda_{k+1} - \bar{\lambda}(x_k, \lambda_k, \mu_k)$ and where the summation is null if $j = 1$. We also have

$$\mu_{k_{i+1}} = \mu_{k_i+j} = \mu_{k_i+1} = \tau\, \mu_{k_i}. \tag{4.6}$$

Now suppose that $j > 1$. Then for the set of iterations $k_i + l, 1 \leq l < j$, either (3.3) or (3.5) must hold and hence the inequality $\|c(x_{k_i+l})\| \leq \gamma_2 \eta_{k_i+l}$ must hold. From this last inequality, (4.6) and the recursive definition of η_k, we must also have

$$\|c(x_{k_i+l})\| \leq \gamma_2 \eta_{k_i+l} = \gamma_2 \eta_{k_i+1} \mu_{k_i+1}^{\beta_\eta(l-1)} = \gamma_2 \mu_{k_{i+1}}^{\beta_\eta(l-1)+\alpha_\eta}. \tag{4.7}$$

On the other hand, from (3.6), (4.6) and the recursive definition of ω_k, we have

$$\|e_{k_i+l}\| \leq \omega_{k_i+l} = \omega_{k_i+1} \mu_{k_i+1}^{l-1} = \mu_{k_{i+1}}^{l}. \tag{4.8}$$

Combining equations (4.4) to (4.8), we obtain the bound

$$\begin{aligned} \|\lambda_{k_i+j}\| &\leq \|\lambda_{k_i}\| + \textstyle\sum_{l=1}^{j-1} \left[\|c(x_{k_i+l})\|/\mu_{k_i+l} + \|e_{k_i+l}\|\right] \\ &\leq \|\lambda_{k_i}\| + \gamma_2 \mu_{k_{i+1}}^{\alpha_\eta - 1} \textstyle\sum_{l=1}^{j-1} \mu_{k_{i+1}}^{\beta_\eta(l-1)} + \textstyle\sum_{l=1}^{j-1} \mu_{k_{i+1}}^{l} \\ &\leq \|\lambda_{k_i}\| + \gamma_2 \mu_{k_{i+1}}^{\alpha_\eta - 1}/(1 - \mu_{k_{i+1}}^{\beta_\eta}) + \mu_{k_{i+1}}/(1 - \mu_{k_{i+1}}) \\ &\leq \|\lambda_{k_i}\| + 2\gamma_2 \mu_{k_{i+1}}^{\alpha_\eta - 1} + 2\mu_{k_{i+1}}. \end{aligned}$$

Thus we obtain that

$$\begin{aligned} \mu_{k_i+j}\|\lambda_{k_i+j}\| &\leq \tau\mu_{k_i}\|\lambda_{k_i}\| + 2\gamma_2\mu_{k_{i+1}}^{\alpha_\eta} + 2\mu_{k_{i+1}}^2 \\ &\leq \tau\mu_{k_i}\|\lambda_{k_i}\| + 2(\gamma_2+1)\mu_{k_{i+1}}^{\alpha_\eta}, \end{aligned} \tag{4.9}$$

where we used (4.6), $0 < \alpha_\eta < 1$ and $\mu_{k_{i+1}} < 1$ to derive the last inequality. Equation (4.9) is also satisfied when $j = 1$ as (3.8) and (4.6) give $\mu_{k_i+1}\|\lambda_{k_i+1}\| = \tau\mu_{k_i}\|\lambda_{k_i}\|$. Hence, from (4.9),

$$\mu_{k_{i+1}}\|\lambda_{k_{i+1}}\| \leq \tau\mu_{k_i}\|\lambda_{k_i}\| + 2(\gamma_2+1)\mu_{k_{i+1}}^{\alpha_\eta}. \tag{4.10}$$

Equation (4.10) then gives that $\mu_{k_i}\|\lambda_{k_i}\|$ converges to zero as k increases. For, if we define

$$\delta_i \stackrel{\text{def}}{=} \mu_{k_i}\|\lambda_{k_i}\| \quad \text{and} \quad \rho_i \stackrel{\text{def}}{=} 2(\gamma_2+1)\mu_{k_i}^{\alpha_\eta}, \tag{4.11}$$

equations (4.6), (4.10) and (4.11) give that

$$\delta_{i+1} \leq \tau\delta_i + \tau^{\alpha_\eta}\rho_i \quad \text{and} \quad \rho_{i+1} = \tau^{\alpha_\eta}\rho_i \tag{4.12}$$

and hence that

$$0 \leq \delta_i \leq \tau^i\delta_0 + (\tau^{\alpha_\eta})^i \sum_{l=0}^{i-1}(\tau^{1-\alpha_\eta})^l\rho_0. \tag{4.13}$$

Since $0 < \tau < 1$ and $0 < \alpha_\eta < 1$, the sum in (4.13) can be bounded to give

$$0 \leq \delta_i \leq \tau^i\delta_0 + (\tau^{\alpha_\eta})^i\rho_0/(1-\tau^{1-\alpha_\eta}). \tag{4.14}$$

But both δ_0 and ρ_0 are finite. Thus, as i increases, δ_i converges to zero. Moreover equation (4.11) implies that ρ_i converges to zero. Therefore, as the right-hand-side of (4.9) converges to zero, the whole sequence $\mu_k\|\lambda_k\|$ converges to zero and the truth of the lemma is established. □

The next theorem gives the desired global convergence property of Algorithm ALINF, analogously to Theorem 4.6 in [2].

Theorem 4.3 *Assume that AS1 holds. Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm ALINF of §3 for which AS2 and AS3 hold and let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*. Finally, let $\lambda_* = \lambda(x_*)$. Then conclusions (i), (ii) and (iii) of Lemma 4.1 hold.*

Again the proof of Theorem 4.3 is detailed in the appendix. It can be easily derived from the proof of Theorem 4.6 in [2], where, in the case where μ_k is supposed to be bounded away from zero, we can use the fact that Step 3a implies that $\|c(x_k)\| \leq \gamma_2\eta_k$ is always satisfied for k large enough, such that $c(x_k)$ converges to zero, by (3.9).

We end this section by briefly analysing second-order conditions. By imposing the following additional assumption, we can guarantee that Algorithm ALINF converges to an isolated local solution.

AS4: Suppose that x_k, λ_k and μ_k are generated by Algorithm ALINF of §3 and that the sequence $\{x_k\}$ converges to x_* for $k \in \mathcal{K}$. Then, we assume that $\nabla_{xx}\Phi(x_k, \lambda_k, \mu_k)$ is uniformly positive definite (that is, its smallest eigenvalue is uniformly bounded away from zero) for all $k \in \mathcal{K}$ sufficiently large.

The following theorem is inspired from Theorem 6.1 in [2].

Theorem 4.4 *Assume that AS1 holds. Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm ALINF of §3 for which AS2 and AS3 hold and let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*. Under Assumption AS4, the iterates x_k, $k \in \mathcal{K}$, converge to an isolated local solution of (1.1).*

The proof is given in the appendix.

4.2 Asymptotic convergence analysis

We now analyse the asymptotic convergence of Algorithm ALINF and show first that the penalty parameter is bounded away from zero. We require some additional assumptions.

AS5: The second Fréchet–derivatives of the functions $f(x)$ and $c(x)$ are Lipschitz continuous at any limit point x_* of the sequence of iterates $\{x_k\}$.

AS6: Suppose that (x_*, λ_*) is a Karush-Kuhn-Tucker point for problem (1.1). We assume that the operator

$$\begin{pmatrix} \nabla_{xx}\ell(x_*, \lambda_*) & c'(x_*)^* \\ c'(x_*) & 0 \end{pmatrix}$$

has a continuous inverse.

Note that AS6 implies AS3.

From the adaptation of Lemma 5.3, (i), in [2] to our framework, we get the following results.

Lemma 4.5 *Assume that AS1 and AS2 hold. Let $\{x_k\}$, $k \in \mathcal{K}$, be a convergent subsequence of iterates produced by Algorithm ALINF, whose limit point is x_* with corresponding Lagrange multiplier λ_*. Assume that AS5 and AS6 hold at x_*. Assume furthermore that μ_k tends to zero as k increases. Then there are positive constants $\bar{\mu} < 1$, κ_3, κ_4, κ_5, κ_6 and an integer k_1 such that, if $\mu_{k_1} \leq \bar{\mu}$, then*

$$\|x_k - x_*\| \leq \kappa_3\omega_k + \kappa_4\mu_k\|\lambda_k - \lambda_*\|, \tag{4.15}$$

$$\|\bar{\lambda}(x_k, \lambda_k, \mu_k) - \lambda_*\| \leq \kappa_5\omega_k + \kappa_6\mu_k\|\lambda_k - \lambda_*\|, \tag{4.16}$$

and

$$\|c(x_k)\| \leq \kappa_5\omega_k\mu_k + \mu_k(1 + \kappa_6\mu_k)\|\lambda_k - \lambda_*\|, \tag{4.17}$$

for all $k \geq k_1$, $k \in \mathcal{K}$.

The proof of Lemma 4.5 is independent of the two relaxations intoduced in Algorithm ALINF, and is detailed in the appendix.

We now restrict our attention to the case where the whole sequence of iterates converges to x_*, making Assumption AS2 unnecessary (see [3] for a motivation of this additional assumption). We then show that, if the penalty parameter μ_k converges to zero, the Lagrange multiplier estimate λ_k converges to its true value λ_* (as in Lemma 5.4 in [2]).

Lemma 4.6 *Assume that AS1 holds. Assume that $\{x_k\}$, the sequence of iterates generated by Algorithm ALINF, converges to x_* at which AS6 holds, and with corresponding Lagrange multiplier λ_*. Then, if μ_k tends to zero, the sequence λ_k converges to λ_*.*

Proof. Recall that AS6 implies AS3 and therefore that our assumptions are sufficient to apply Theorem 4.3.

The result is obvious if Step 3a is executed infinitely often because each time this step is executed, $\lambda_{k+1} = \bar{\lambda}(x_k, \lambda_k, \mu_k) + e_k$, with $e_k \stackrel{\text{def}}{=} \lambda_{k+1} - \bar{\lambda}(x_k, \lambda_k, \mu_k)$ satisfying $\|e_k\| \le \omega_k$. This, together with inequality (4.1), guarantees that λ_k converges to λ_*. Suppose, therefore, that Step 3a is not executed infinitely often. Then $\|\lambda_k - \lambda_*\|$ will remain fixed for all $k \ge k_2$, for some $k_2 > 0$, as Step 3b is executed for each remaining iteration. But (4.3) then implies that $\|c(x_k)\| \le \kappa_{13}\mu_k$ for some constant $\kappa_{13} > 0$ and for all $k \ge k_3 \ge k_2$. As μ_k tends to zero and $0 < \alpha_\eta < 1$, $\kappa_{13}\mu_k \le \gamma_1 \mu_k^{\alpha_\eta} = \gamma_1 \eta_k$ for all k sufficiently large for which Step 3b is executed. But then inequality (3.3) must be satisfied for some $k \ge k_3$, which is impossible, since this would imply that Step 3a is again executed. Hence Step 3a must be executed infinitely often. $\square$

We now consider the behaviour of the penalty parameter μ_k and show that it is bounded away from zero, avoiding to the Hessian of the augmented Lagrangian to become increasingly ill-conditioned. The proof of this result follows the lines of Theorem 5.5 in [2].

Theorem 4.7 *Assume AS1 holds and suppose that the sequence of iterates $\{x_k\}$ of Algorithm ALINF converges to x_* with corresponding Lagrange multiplier λ_*, at which AS5 and AS6 hold. Then there is a constant $\mu_{\min} \in (0, 1)$ such that $\mu_k \ge \mu_{\min}$ for all k.*

Proof. Since in any case μ_k is a nonincreasing sequence, we suppose by contradiction that the complete sequence μ_k tends to zero. Then Step 3b must be executed infinitely often. We note that our assumptions are sufficient to apply Theorem 4.3. Furthermore, we may apply Lemma 4.5 to the complete sequence of iterates.

First observe that

$$\mu_k \le \bar{\mu} < 1 \tag{4.18}$$

for all $k \ge k_1$, where $\bar{\mu}$ and k_1 are those of Lemma 4.5. Note that

$$\omega_k \le \mu_k$$

for all $k \ge k_1$. This follows by definition if (3.8) is executed. Otherwise it is a consequence of the fact that μ_k is unchanged while ω_k is reduced, when (3.7) occurs. Let k_4 be the smallest integer k such that, for all $k \ge k_4$,

$$\mu_k^{1-\alpha_\eta} \le \frac{\gamma_1}{2 + \kappa_5}, \tag{4.19}$$

and

$$\mu_k^{1-\beta_\eta} \le \min\left[\frac{1}{\kappa_{14}}, \frac{\gamma_1}{2\kappa_{14}+\kappa_5}\right], \tag{4.20}$$

where $\kappa_{14} = 1 + \kappa_5 + \kappa_6$. Note that (4.18), $0 < \beta_\eta < 1$ and (4.20) imply that

$$\mu_k \le \mu_k^{1-\beta_\eta} \le \frac{1}{\kappa_{14}} \le \frac{1}{\kappa_6} \tag{4.21}$$

for all $k \ge \max(k_1, k_4)$. Furthermore, let k_5 be such that

$$\|\lambda_k - \lambda_*\| \le 1 \tag{4.22}$$

for all $k \ge k_5$, which is possible because of Lemma 4.6. Now define $k_6 = \max(k_1, k_4, k_5)$, let Γ be the set $\{k \mid (3.8) \text{ is executed at iteration } k-1 \text{ and } k \ge k_6\}$ and let k_0 be the smallest element of Γ. By the assumption that μ_k tends to zero, Γ has an infinite number of elements.

By the definition of Γ, for iteration k_0, $\omega_{k_0} = \mu_{k_0}$ and $\eta_{k_0} = \mu_{k_0}^{\alpha_\eta}$. Then inequality (4.17) gives that,

$$\begin{array}{rcll} \|c(x_{k_0})\| & \le & \kappa_5\omega_{k_0}\mu_{k_0} + \mu_{k_0}(1+\kappa_6\mu_{k_0})\|\lambda_{k_0} - \lambda_*\| & \\ & \le & \kappa_5\omega_{k_0}\mu_{k_0} + 2\mu_{k_0}\|\lambda_{k_0} - \lambda_*\| & \text{(from (4.21))} \\ & \le & (2+\kappa_5\mu_{k_0})\mu_{k_0} & \text{(from (4.22))} \\ & \le & (2+\kappa_5)\mu_{k_0} & \text{(from (4.18))} \\ & \le & \gamma_1\mu_{k_0}^{\alpha_\eta} = \gamma_1\eta_{k_0} & \text{(from (4.19)).} \end{array} \tag{4.23}$$

As a consequence of this inequality, Step 3a will be executed with $\lambda_{k_0+1} = \bar\lambda(x_{k_0}, \lambda_{k_0}, \mu_{k_0}) + e_{k_0}$, where $\|e_{k_0}\| \le \omega_{k_0}$ by (3.6). Inequality (4.16) together with (4.22) then guarantee that

$$\|\lambda_{k_0+1} - \lambda_*\| \le (\kappa_5+1)\omega_{k_0} + \kappa_6\mu_{k_0}\|\lambda_{k_0} - \lambda_*\| \le \kappa_{14}\mu_{k_0}. \tag{4.24}$$

We shall now make use of an inductive proof. Assume that Step 3a is executed for iterations $k_0 + i$, $(0 \le i \le t)$, and that

$$\|\lambda_{k_0+i+1} - \lambda_*\| \le \kappa_{14}\mu_{k_0}^{1+\beta_\eta i}. \tag{4.25}$$

Inequalities (4.23) and (4.24) show that this is true for $t = 0$. We aim to show that the same is true for $i = t+1$. Our assumption that Step 3a is executed gives that, for iteration $k_0 + t + 1$, $\mu_{k_0+t+1} = \mu_{k_0}$, $\omega_{k_0+t+1} = \mu_{k_0}^{t+2}$, and $\eta_{k_0+t+1} = \mu_{k_0}^{\beta_\eta(t+1)+\alpha_\eta}$. Then, inequality (4.17) yields that

$$\begin{array}{rcll} \|c(x_{k_0+t+1})\| & \le & \kappa_5\omega_{k_0+t+1}\mu_{k_0+t+1} & \\ & & +\mu_{k_0+t+1}(1+\kappa_6\mu_{k_0+t+1})\|\lambda_{k_0+t+1} - \lambda_*\| & \\ & \le & \kappa_5\omega_{k_0+t+1}\mu_{k_0+t+1} + 2\mu_{k_0+t+1}\|\lambda_{k_0+t+1} - \lambda_*\| & \text{(from (4.21))} \\ & \le & \kappa_5\mu_{k_0}^{t+3} + 2\kappa_{14}\mu_{k_0}\mu_{k_0}^{1+\beta_\eta t} & \text{(from (4.25))} \\ & \le & \kappa_5\mu_{k_0}^{\alpha_\eta+\beta_\eta(t+1)+1} + 2\kappa_{14}\mu_{k_0}\mu_{k_0}^{\alpha_\eta+\beta_\eta t} & \text{(since } 0 < \alpha_\eta, \beta_\eta < 1\text{)} \\ & \le & (2\kappa_{14}+\kappa_5)\mu_{k_0}^{1-\beta_\eta}\mu_{k_0}^{\beta_\eta(t+1)+\alpha_\eta} & \text{(from (4.21))} \\ & \le & \gamma_1\mu_{k_0}^{\beta_\eta(t+1)+\alpha_\eta} = \gamma_1\eta_{k_0+t+1} & \text{(from (4.20)).} \end{array}$$

Hence Step 3a will again be executed with

$$\lambda_{k_0+t+2} = \bar{\lambda}(x_{k_0+t+1}, \lambda_{k_0+t+1}, \mu_{k_0+t+1}) + e_{k_0+t+1},$$

where $\|e_{k_0+t+1}\| \leq \omega_{k_0+t+1}$ by (3.6). Inequality (4.16) then implies that

$$\begin{aligned}
\|\lambda_{k_0+t+2} - \lambda_*\| &\leq (\kappa_5 + 1)\omega_{k_0+t+1} + \kappa_6 \mu_{k_0+t+1} \|\lambda_{k_0+t+1} - \lambda_*\| \\
&\leq (\kappa_5 + 1)\mu_{k_0}^{t+2} + \kappa_6 \kappa_{14} \mu_{k_0} \mu_{k_0}^{1+\beta_\eta t} && \text{(from (4.25))} \\
&\leq (\kappa_5 + 1)\mu_{k_0}^{1+\beta_\eta(t+1)} + \kappa_6 \kappa_{14} \mu_{k_0} \mu_{k_0}^{1+\beta_\eta t} \\
&= (\kappa_5 + 1 + \kappa_6 \kappa_{14} \mu_{k_0}^{1-\beta_\eta}) \mu_{k_0}^{1+\beta_\eta(t+1)} \\
&\leq (\kappa_5 + 1 + \kappa_6) \mu_{k_0}^{1+\beta_\eta(t+1)} && \text{(from (4.20))} \\
&\leq \kappa_{14} \mu_{k_0}^{1+\beta_\eta(t+1)},
\end{aligned}$$

which establishes (4.25) for $i = t + 1$. Thus Step 3a is executed for all iterations $k \geq k_0$. But this implies that Γ is finite, which contradicts the assumption that Step 3b is executed infinitely often. Hence the theorem is proved. $\Box$

As in [2], we examine the rate of convergence of our algorithm. Since we now allow an error on the first-order update of the Lagrange multiplier estimate in Algorithm ALINF, we have to distinguish between the sequences $\bar{\lambda}(x_k, \lambda_k, \mu_k)$ and λ_k.

Theorem 4.8 *Under the assumptions of Theorem 4.7, the iterates x_k and the Lagrange multipliers $\bar{\lambda}(x_k, \lambda_k, \mu_k)$ and λ_k of Algorithm ALINF are at least R-linearly convergent with R-factor at most $\mu_{\min}^{\beta_\eta}$, where $\mu_{\min}$ is the smallest value of the penalty parameter generated by the algorithm.*

The proof is given in the appendix for completeness, and makes use of the fact that Step 3a implies that $\|c(x_k)\| \leq \gamma_2 \eta_k$ is always satisfied for k sufficiently large. The rate of convergence of the sequence λ_k is a direct consequence of (3.6).

5 Approximation of optimization problems

5.1 Approximation Scheme

In this section we consider an approximation of an optimization problem posed in an infinite-dimensional space. More precisely, we replace the original problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad c(x) = 0, \quad x \in X,$$

by a sequence of problems

$$\text{minimize} \quad f_n(x) \quad \text{subject to} \quad c_n(x) = 0, \quad x \in X_n,$$

where the functions and mappings are defined as follows

$$f_n : X_n \to I\!R, \qquad c_n : X_n \to Y_n,$$

with sequences of Hilbert spaces X_n and Y_n, for $n \in I\!N$, over the reals. This occurs for example by proper discretization of optimal control problems where n denotes the parameter describing the discretization level.

We denote by $\Phi_n : X_n \times Y_n \times I\!R \to I\!R$ the augmented Lagrangian function for these problems

$$\Phi_n(x,\lambda,\mu) = f_n(x) + < \lambda, c_n(x) >_{Y_n} + \frac{1}{2\mu}\|c_n(x)\|^2_{Y_n}, \quad x \in X_n,\ \lambda \in Y_n. \tag{5.1}$$

On the functions we first make the following extension of the original Assumption AS1.

AS1$_n$: The functions f, f_n and c, c_n are twice continuously Fréchet-differentiable for all $n \in I\!N$.

We assume also that the spaces X_n and Y_n are nested

$$X_1 \subseteq \ldots \subseteq X_n \subseteq X_{n+1} \subseteq \ldots \subseteq X \qquad \text{and} \qquad Y_1 \subseteq \ldots \subseteq Y_n \subseteq Y_{n+1} \subseteq \ldots \subseteq Y.$$

For the inner products and norms in these spaces we choose

$$< \cdot,\cdot >_{X_n} = < \cdot,\cdot >_X, \quad < \cdot,\cdot >_{Y_n} = < \cdot,\cdot >_Y, \quad \|\cdot\|_{X_n} = \|\cdot\|_X, \quad \|\cdot\|_{Y_n} = \|\cdot\|_Y.$$

We further make the following assumption on the functions for the refinement of the discretization.

AS7: There exists $n^\star \in I\!N$ such that

a. for all $x \in X_{n^\star}$, there exists a sequence of positive constants $\varepsilon_{c,n}$ for which

$$\lim_{n\to\infty} \varepsilon_{c,n} = 0 \ \text{ and } \ \|c_n(x) - c(x)\|_Y \le \varepsilon_{c,n} \ \text{ for } \ n \ge n^\star, \ \text{ and}$$

b. for all $x \in X_{n^\star}$, $\lambda \in Y_{n^\star}$ and $\mu > 0$, there exists a sequence of positive constants $\varepsilon_{\Phi,n}$ for which

$$\lim_{n\to\infty} \varepsilon_{\Phi,n} = 0 \ \text{ and } \ \|\nabla_x \Phi_n(x,\lambda,\mu) - \nabla_x \Phi(x,\lambda,\mu)\|_X \le \varepsilon_{\Phi,n} \ \text{ for } \ n \ge n^\star.$$

The goal for considering the following algorithm is not only to adapt the penalty parameter and tolerance level for the solution of the unconstrained subproblems but also to incorporate the discretization aspect. It is not of much value to solve a subproblem very precisely at the early stage of an iteration. Therefore one would like to have a measure with which one can decide on the refinement of the discretization when necessary. Step 4 in the algorithm below includes such a decision.

Algorithm ALDISCR

Step 0 [Initialization]. Let $\{\varepsilon_{c,n}\}_{n\in\mathbb{N}}$ and $\{\varepsilon_{\Phi,n}\}_{n\in\mathbb{N}}$ be sequences of positive constants converging to zero. An initial discretization level n_0, a Lagrange multiplier estimate $\lambda_0 \in Y_{n_0}$ and a positive penalty parameter $\mu_0 < 1$ are given. The positive constants $\omega_* < 1$, $\eta_* < 1$, $\tau < 1$, $\alpha < 1$, $\alpha_\eta < 1$, and $\beta_\eta < 1$ are specified. Set $\omega_0 = \mu_0$, $\eta_0 = \mu_0^{\alpha_\eta}$, and $k = 0$.

Step 1 [Inner iteration]. Find $x_k \in X_{n_k}$ such that

$$\|\nabla_x \Phi_{n_k}(x_k, \lambda_k, \mu_k)\|_X \leq \omega_k/2 \tag{5.2}$$

holds.

Step 2 [Test for convergence]. If $\omega_k \leq \omega_*$, $\|c_{n_k}(x_k)\|_Y \leq \eta_*/2$ and $\varepsilon_{c,n_k} \leq \eta_*/2$, stop.

Step 3 [Updates]. If

$$\|c_{n_k}(x_k)\|_Y \leq \eta_k, \tag{5.3}$$

execute Step 3a. If

$$\|c_{n_k}(x_k)\|_Y > \eta_k, \tag{5.4}$$

execute Step 3b.

Step 3a [Update Lagrange multiplier estimate]. Choose

$$\lambda_{k+1} = \lambda_k + \frac{1}{\mu_k} c_{n_k}(x_k) \tag{5.5}$$

and set

$$\begin{array}{rcl} \mu_{k+1} & = & \mu_k, \\ \omega_{k+1} & = & \omega_k \mu_{k+1}, \\ \eta_{k+1} & = & \eta_k \mu_{k+1}^{\beta_\eta}. \end{array} \tag{5.6}$$

Step 3b [Reduce the penalty parameter]. Set

$$\begin{array}{rcl} \lambda_{k+1} & = & \lambda_k, \\ \mu_{k+1} & = & \tau \mu_k, \\ \omega_{k+1} & = & \mu_{k+1}, \\ \eta_{k+1} & = & \mu_{k+1}^{\alpha_\eta}. \end{array} \tag{5.7}$$

Step 4 [Refinement]. Choose $n_{k+1} \geq n_k$ such that

$$\begin{array}{rcl} \varepsilon_{c,n_{k+1}} & < & \min\{\alpha\eta_{k+1}, \mu_{k+1}\omega_{k+1}\}, \\ \varepsilon_{\Phi,n_{k+1}} & \leq & \dfrac{\omega_{k+1}}{2}. \end{array} \tag{5.8}$$

Increment k by one and go to Step 1.

We now make some comments on Algorithm ALDISCR.

- The quantity ω_k is chosen as a termination criterion in Step 2 of Algorithm ALDISCR because it is also an upper bound on $\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X$, as proved in Lemma 5.1.
- Assumption AS7 and Step 2 of Algorithm ALDISCR imply that $\|c(x_k)\|_Y \leq \eta_*$. Indeed, by the first part of Assumption AS7, we have that

$$\|c(x_k)\|_Y - \|c_{n_k}(x_k)\|_Y \leq |\|c(x_k)\|_Y - \|c_{n_k}(x_k)\|_Y| \leq \|c(x_k) - c_{n_k}(x_k)\|_Y \leq \varepsilon_{c,n_k},$$

  which implies that

$$\|c(x_k)\|_Y \leq \varepsilon_{c,n_k} + \|c_{n_k}(x_k)\|_Y \leq \eta_*,$$

  by the two last conditions in Step 2.
- The choice of the constant $\alpha < 1$ in Step 0 specifies the parameters γ_1 and γ_2 for the decision rule of Step 3 in Algorithm ALINF, as proved in Lemma 5.2.

The algorithm is well defined since the subspaces are nested and $\lambda_{k+1} \in Y_{n_{k+1}}$ by (5.5), while $x_{k+1} \in X_{n_{k+1}}$ by Step 1. In addition we know that all iterates x_k lie in X. This indicates that one would like to have a convergence result which is formulated in the original space X. Therefore we interpret the x_k as iterates of the infinite-dimensional Algorithm ALINF. To prove this assertion we verify the different steps of the algorithm.

Lemma 5.1 *Let Assumptions AS1$_n$ and AS7 hold. For each iterate x_k of Algorithm ALDISCR we have*

$$\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X \leq \omega_k \,.$$

Moreover, the sequence of Lagrange multipliers λ_k of Algorithm ALDISCR satisfies

$$\|\lambda_{k+1} - \bar{\lambda}(x_k, \lambda_k, \mu_k)\|_Y \leq \omega_k.$$

Proof. We make the following estimate using (5.2), Assumption AS7 and (5.8)

$$\begin{aligned} \|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X &\leq \|\nabla_x \Phi_{n_k}(x_k, \lambda_k, \mu_k)\|_X + \|\nabla_x \Phi(x_k, \lambda_k, \mu_k) - \nabla_x \Phi_{n_k}(x_k, \lambda_k, \mu_k)\|_X \\ &\leq \frac{\omega_k}{2} + \varepsilon_{\Phi,n_k} \leq \frac{\omega_k}{2} + \frac{\omega_k}{2} = \omega_k. \end{aligned}$$

To show the estimate on the Lagrange multiplier note that by (2.1), (5.5), Assumption AS7 and (5.8)

$$\|\lambda_{k+1} - \bar{\lambda}(x_k, \lambda_k, \mu_k)\|_Y = \frac{1}{\mu_k} \, \|c_{n_k}(x_k) - c(x_k)\|_Y \leq \frac{1}{\mu_k} \, \varepsilon_{c,n_k} \leq \omega_k.$$

□

In the following lemma we show that if the decision rule of Algorithm ALDISCR is applied, its iterates satisfy also the decision rule of Algorithm ALINF.

Lemma 5.2 *Let Assumption AS7 hold and for* $\alpha \in (0,1)$ *set* $\gamma_1 = 1-\alpha$ *and* $\gamma_2 = 1+\alpha$. *For any iterate* x_k *of Algorithm ALDISCR we have the implication*

$$\begin{array}{llll} \text{If } \|c(x_k)\|_Y & \geq & \gamma_2\eta_k, & \text{then Step 3b of Algorithm ALINF is executed.} \\ \text{If } \|c(x_k)\|_Y & \leq & \gamma_1\eta_k, & \text{then Step 3a of Algorithm ALINF is executed.} \end{array}$$

Proof. By (5.8) we have

$$\varepsilon_{c,n_k} < \alpha\eta_k.$$

Let $\|c(x_k)\|_Y \geq \gamma_2\eta_k$. Then, by Assumption AS7

$$\begin{aligned} \|c_{n_k}(x_k)\|_Y &\geq \|c(x_k)\|_Y - \|c(x_k) - c_{n_k}(x_k)\|_Y \\ &\geq \gamma_2\eta_k - \varepsilon_{c,n_k} > \gamma_2\eta_k - \alpha\eta_k = \eta_k\,. \end{aligned}$$

By Algorithm ALDISCR, Step 3b is then executed which is identical with Step 3b of Algorithm ALINF. On the other hand suppose that $\|c(x_k)\|_Y \leq \gamma_1\eta_k$ is true. Then Assumption AS7 implies

$$\begin{aligned} \|c_{n_k}(x_k)\|_Y &\leq \|c(x_k)\|_Y + \|c(x_k) - c_{n_k}(x_k)\|_Y \\ &\leq \gamma_1\eta_k + \varepsilon_{c,n_k} < \gamma_1\eta_k + \alpha\eta_k = \eta_k\,. \end{aligned}$$

By Algorithm ALDISCR, Step 3a is executed and the Lagrange multiplier λ_k is updated according to the rule (5.5). In the previous Lemma 5.1 it was shown that this multiplier also satisfies the relation (3.6). Since the update formulas for μ, ω and η are identical in both algorithms, we conclude that Step 3a of Algorithm ALINF is executed if $\|c(x_k)\|_Y \leq \gamma_1\eta_k$. □

We can summarize these lemmas in the following theorem.

Theorem 5.3 *Let* $AS1_n$ *and AS7 hold. Then the sequence* $\{x_k\}$ *generated by Algorithm ALDISCR can be considered as a sequence generated by Algorithm ALINF. Furthermore, if* x_* *is a limit point of this sequence for which AS2 and AS3 hold, if* $\mathcal{K}$ *is the set of indices of an infinite subsequence of the* x_k *whose limit is* x_* *and if we set* $\lambda_* = \lambda(x_*)$*, then conclusions (i), (ii) and (iii) of Lemma 4.1 hold.*

Theorem 4.8 of §4.2 about the rate of convergence applies here too.

Theorem 5.4 *If Assumptions* $AS1_n$*, AS7 and the assumptions of Theorem 4.7 hold, the iterates* x_k *and the Lagrange multipliers* $\bar\lambda(x_k, \lambda_k, \mu_k)$ *and* λ_k *of Algorithm ALDISCR are at least R-linearly convergent with R-factor at most* $\mu_{\min}^{\beta_\eta}$*, where* $\mu_{\min}$ *is the smallest value of the penalty parameter generated by the algorithm.*

Note finally that one can show that the penalty parameter is bounded away from zero and that the Lagrange multipliers are bounded using the results in Theorems 4.7 and 4.8. Therefore, the reduction in the discretization parameters ε in (5.8) is determined by the quantity ω_k.

5.2 Application to Optimal Control

In addition to the sequences of finite-dimensional Hilbert spaces $X_n \subseteq X$ and $Y_n \subseteq Y$, $n \in I\!N$, we introduce restrictions and prolongations as linear bounded maps defined on the subspaces X_n and Y_n

$$r_n^X : X \to X_n, \quad r_n^Y : Y \to Y_n, \qquad p_n^X : X_n \to X, \quad p_n^Y : Y_n \to Y.$$

We have the following result.

Lemma 5.5 *Let X be a Hilbert space and $X_n \subseteq X$. If the prolongation $p_n^X : X_n \to X$ is given by the embedding operator $p_n^X = \iota_n^X$, then the dual operator $(p_n^X)^* : X \to X_n$ as a restriction operator is given by the orthogonal projection $(p_n^X)^* = \pi_n^X$ onto X_n.*

This follows easily from the definition of the dual map. We thus define the restrictions and prolongations as

$$p_n^X = \iota_n^X, \quad p_n^Y = \iota_n^Y, \qquad r_n^X = \pi_n^X, \quad r_n^Y = \pi_n^Y.$$

We make the following assumption on the selection of the subspaces.

AS8: The Hilbert spaces $X_n \subseteq X$ and $Y_n \subseteq Y$, $n \in I\!N$, are chosen such that the orthogonal projections $\pi_n^X \in \mathcal{L}(X, X_n)$ and $\pi_n^Y \in \mathcal{L}(Y, Y_n)$ satisfy, for each $x \in X$ and $y \in Y$, respectively,

$$\lim_{n\to\infty} \|\pi_n^X x - x\|_X = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \|\pi_n^Y y - y\|_Y = 0.$$

The discretized functions and mappings are then defined as follows

AS9: For all $n \in I\!N$

$$f_n = f \circ \iota_n^X : X_n \to I\!R, \qquad c_n = \pi_n^Y \circ c \circ \iota_n^X : X_n \to Y_n.$$

We now show that Assumptions AS1$_n$ and AS7 are satisfied under the above setting. Obviously the differentiability assumptions for f_n and c_n in AS1$_n$ directly follow from Assumptions AS8 and AS9.

Lemma 5.6 *Let f and c satisfy Assumption AS1 and let AS8 and AS9 hold. Then f_n and c_n defined in Assumption AS9 satisfy the Assumption AS1$_n$, for all $n \in I\!N$.*

With respect to AS7, we first show in the next statement the consistency of the approximations of c in form of the pointwise convergence of the discretized constraints to the original constraint.

Lemma 5.7 *Let c satisfy AS1, let Assumption AS8 hold and let c_n, $n \in I\!N$, satisfy AS9. Then for any $n \in I\!N$ we have that for all $x \in X_n$*

$$\lim_{n\to\infty} \|c_n(x) - c(x)\|_Y = 0.$$

Proof. For fixed $x \in X_n$, we have

$$\|c_n(x) - c(x)\|_Y = \|\pi_n^Y c(\iota_n^X x) - c(x)\|_Y = \|\pi_n^Y c(x) - c(x)\|_Y,$$

which proves the lemma, by Assumption AS8. □

We now turn to the second part of AS7. The augmented Lagrangian function $\Phi_n : X_n \times Y_n \times I\!R \to I\!R$ becomes, with the definition of f_n and c_n given in Assumption AS9,

$$\begin{aligned}
\Phi_n(x, \lambda, \mu) &= f_n(x) + < \lambda, c_n(x) >_{Y_n} + \frac{1}{2\mu}\|c_n(x)\|_{Y_n}^2 \\
&= f(\iota_n^X x) + < \iota_n^Y \lambda, c(\iota_n^X x) >_Y + \frac{1}{2\mu}\|\pi_n^Y c(\iota_n^X x)\|_{Y_n}^2 \\
&= \ell(\iota_n^X x, \iota_n^Y \lambda) + \frac{1}{2\mu}\|\pi_n^Y c(\iota_n^X x)\|_{Y_n}^2.
\end{aligned}$$

The gradient with respect to $x \in X_n$ of the augmented Lagrangian function Φ_n is computed in the following Lemma.

Lemma 5.8 *The gradient of $\Phi_n(x, \lambda, \mu)$ with respect to $x \in X_n$ is given by*

$$\begin{aligned}
\nabla_x \Phi_n(x, \lambda, \mu) &= \pi_n^X \nabla_x f(\iota_n^X x) + \pi_n^X c'(\iota_n^X x)^* \iota_n^Y \lambda + \frac{1}{\mu}\pi_n^X c'(\iota_n^X x)^* \iota_n^Y \pi_n^Y c(\iota_n^X x) \\
&= \pi_n^X \nabla_x \ell(\iota_n^X x, \iota_n^Y \lambda) + \frac{1}{\mu}\pi_n^X c'(\iota_n^X x)^* \iota_n^Y \pi_n^Y c(\iota_n^X x).
\end{aligned}$$

Proof. We have the following identities, where differentiation is understood with respect to $x \in X_n$ and where $\lambda \in Y_n$ and $s \in X_n$

$$\begin{aligned}
\Phi_n'(x, \lambda, \mu)s &= < \nabla_x f(\iota_n^X x), \iota_n^X s >_X + < \iota_n^Y \lambda, c'(\iota_n^X x)\iota_n^X s >_Y \\
&\quad + \frac{1}{\mu} < \pi_n^Y c(\iota_n^X x), \pi_n^Y c'(\iota_n^X x)\iota_n^X s >_{Y_n} \\
&= < \pi_n^X \nabla_x f(\iota_n^X x), s >_{X_n} + < \pi_n^X c'(\iota_n^X x)^* \iota_n^Y \lambda, s >_{X_n} \\
&\quad + \frac{1}{\mu} < \pi_n^X c'(\iota_n^X x)^* \iota_n^Y \pi_n^Y c(\iota_n^X x), s >_{X_n},
\end{aligned}$$

which shows the representation of the gradient. □

Next we prove the convergence of the gradients of the augmented Lagrangian.

Lemma 5.9 *Let f and c satisfy AS1, let Assumption AS8 hold and let f_n and c_n, $n \in I\!N$, satisfy AS9. Then for any $n \in I\!N$ we have that for all $x \in X_n$, all $\lambda \in Y_n$ and all $\mu > 0$*

$$\lim_{n \to \infty} \|\nabla_x \Phi_n(x, \lambda, \mu) - \nabla_x \Phi(x, \lambda, \mu)\|_X = 0.$$

Proof. By Lemma 5.8 and the first equality in (2.2), we split

$$\begin{aligned}
\|\nabla_x \Phi_n(x, \lambda, \mu) - \nabla_x \Phi(x, \lambda, \mu)\|_X &\leq \|\pi_n^X \nabla_x \ell(\iota_n^X x, \iota_n^Y \lambda) - \nabla_x \ell(x, \lambda)\|_X \\
&\quad + \frac{1}{\mu}\|\pi_n^X c'(\iota_n^X x)^* \iota_n^Y \pi_n^Y c(\iota_n^X x) - c'(x)^* c(x)\|_X,
\end{aligned}$$

and estimate each part separately. Let $x \in X_n$, $\lambda \in Y_n$ and $\mu > 0$ be fixed. By Assumption AS8, we first have that

$$\lim_{n\to\infty} \|\pi_n^X \nabla_x \ell(\iota_n^X x, \iota_n^Y \lambda) - \nabla_x \ell(x,\lambda)\|_X = \lim_{n\to\infty} \|\pi_n^X \nabla_x \ell(x,\lambda) - \nabla_x \ell(x,\lambda)\|_X = 0.$$

The following inequalities

$$\begin{aligned}
\|\pi_n^X c'(\iota_n^X x)^* \iota_n^Y \pi_n^Y c(\iota_n^X x) - c'(x)^* c(x)\|_X &= \|\pi_n^X c'(x)^* \pi_n^Y c(x) - c'(x)^* c(x)\|_X \\
&\leq \|\pi_n^X c'(x)^* \pi_n^Y c(x) - \pi_n^X c'(x)^* c(x)\|_X \\
&\quad + \|\pi_n^X c'(x)^* c(x) - c'(x)^* c(x)\|_X \\
&\leq \|\pi_n^X\| \|c'(x)^*\| \|\pi_n^Y c(x) - c(x)\|_Y \\
&\quad + \|\pi_n^X c'(x)^* c(x) - c'(x)^* c(x)\|_X,
\end{aligned}$$

together with Assumption AS8 and the fact that $\|\pi_n^Y\|_{\mathcal{L}(Y,Y_n)} \leq 1$, imply that

$$\lim_{n\to\infty} \|\pi_n^X c'(\iota_n^X x)^* \iota_n^Y \pi_n^Y c(\iota_n^X x) - c'(x)^* c(x)\|_X = 0.$$

We can deduce the result from the two limits above. □

In conclusion, Lemmas 5.6, 5.7 and 5.9 guarantee that, under Assumptions AS8 and AS9, the convergence results of Theorems 5.3 and 5.4 hold.

6 Conclusion

In this paper, we present a convergence theory for an infinite-dimensional setting of an augmented Lagrangian method to solve an equality constrained optimization problem. This problem is solved by a sequence of discretized problems where in the course of the iteration the level of discretization is refined. The penalty parameter and the Lagrange multiplier updates are based on a set-valued map. This and the fact that the subproblems are solved inexactly allow us to interpret the iterates of the discretized problems as a sequence of iterates in an infinite-dimensional setting. We obtain a convergence statement of the discretized iterates in Hilbert space. The assumptions are explained for a discretization scheme with nested subspaces. Numerical experiments are part of future work.

References

[1] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, London, 1982.

[2] A. R. Conn, N. I. M. Gould, A. Sartenaer, and Ph. L. Toint. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. *SIAM Journal on Optimization*, 6(3):674–703, 1996.

[3] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572, 1991.

[4] K. Ito and K. Kunisch. The augmented Lagrangian method for parameter estimation in elliptic systems. *SIAM Journal on Control and Optimization*, 28:113–136, 1990.

[5] K. Ito and K. Kunisch. Augmented Lagrangian-SQP-Methods in Hilbert Spaces and Application to Control in the Coefficients Problems. *SIAM Journal on Optimization*, 6(1):96–125, 1996.

[6] E. W. Sachs and S. Volkwein. Augmented Lagrange-SQP methods with Lipschitz-continuous Lagrange multiplier updates. *SIAM Journal on Numerical Analysis*, 40:233–253, 2002.

[7] S. Volkwein. Mesh Independence of an Augmented Lagrangian-SQP Method in Hilbert Spaces and Control Problems for the Burgers Equation. PhD Thesis, Technische Universität Berlin, 1997.

Appendix

A Proof of Lemma 4.1

As a consequence of AS1–AS3, we have that for $k \in \mathcal{K}$ sufficiently large, $c'(x_k)^+$ exists, is bounded and converges to $c'(x_*)^+$. Thus, we may write

$$\|(c'(x_k)^+)^*\| \le \kappa_1 \tag{A.1}$$

for some constant $\kappa_1 > 0$. The inner iteration termination criterion (3.2) gives that

$$\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\| = \|\nabla_x f(x_k) + c'(x_k)^* \bar{\lambda}_k\| \le \omega_k \tag{A.2}$$

for all $k \in \mathcal{K}$, where $\bar{\lambda}_k \stackrel{\text{def}}{=} \bar{\lambda}(x_k, \lambda_k, \mu_k)$. By Assumptions AS1, AS2, AS3 and (2.3), $\lambda(x)$ is bounded for all x in a neighbourhood of x_*. Thus we may deduce from (2.3), (A.1) and (A.2) that

$$\begin{aligned} \|\bar{\lambda}_k - \lambda(x_k)\| &= \|(c'(x_k)^+)^* \nabla_x f(x_k) + \bar{\lambda}_k\| \\ &= \|(c'(x_k)^+)^* (\nabla_x f(x_k) + c'(x_k)^* \bar{\lambda}_k)\| \\ &\le \|(c'(x_k)^+)^*\| \omega_k \\ &\le \kappa_1 \omega_k. \end{aligned} \tag{A.3}$$

Moreover, since $c'(x)$ is Lipschitz-continuous in a neighborhood of x_*, $c'(x)^+$ is also Lipschitz-continuous and

$$\|\lambda(x_k) - \lambda_*\| \le \kappa_2 \|x_k - x_*\| \tag{A.4}$$

for all $k \in \mathcal{K}$ sufficiently large and for some constant $\kappa_2 > 0$, which implies inequality (4.2). We then have that $\lambda(x_k)$ converges to λ_*. Combining (A.3) and (A.4) we obtain

$$\|\bar{\lambda}_k - \lambda_*\| \le \|\bar{\lambda}_k - \lambda(x_k)\| + \|\lambda(x_k) - \lambda_*\| \le \kappa_1 \omega_k + \kappa_2 \|x_k - x_*\|, \tag{A.5}$$

which gives the required inequality (4.1). Then, since by assumption ω_k tends to zero as k increases, (A.5) implies that $\bar{\lambda}_k$ converges to λ_* and therefore $\nabla_x \Phi(x_k, \lambda_k, \mu_k)$ converges to $\nabla_x \ell(x_*, \lambda_*)$. Furthermore, multiplying (2.1) by μ_k, we obtain

$$c(x_k) = \mu_k((\bar{\lambda}_k - \lambda_*) + (\lambda_* - \lambda_k)). \tag{A.6}$$

Taking norms of (A.6) and using (A.5), we derive (4.3).

Now suppose that

$$c(x_*) = 0. \tag{A.7}$$

The convergence of $\nabla_x \Phi(x_k, \lambda_k, \mu_k)$ to $\nabla_x \ell(x_*, \lambda_*)$, (3.2) and the convergence of ω_k to zero give that

$$\nabla_x f(x_*) + c'(x_*)^* \lambda_* = 0.$$

This last equation and (A.7) show that x_* is a Karush-Kuhn-Tucker point and λ_* is the corresponding Lagrange multiplier. Moreover (4.1) and (4.2) ensure the convergence of the sequences $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ and $\{\lambda(x_k)\}$ to λ_* for $k \in \mathcal{K}$. Hence the lemma is proved. □

B Proof of Theorem 4.3

Our assumptions are sufficient to reach the conclusions of part (i) of Lemma 4.1. We now show that $c(x_*) = 0$ and analyse two separate cases.

The first case is when μ_k is bounded away from zero. Hence Step 3a must be executed every iteration for k sufficiently large, implying that $\|c(x_k)\| \leq \gamma_2 \eta_k$ is always satisfied for k large enough. We then deduce from (3.9) that $c(x_k)$ converges to zero.

The second case is when μ_k converges to zero. Then Lemma 4.2 shows that $\mu_k \|(\lambda_k - \lambda_*)\|$ tends to zero. Using this limit and (3.9) in (4.3), we obtain that $c(x_k)$ tends to zero, as desired.

As a consequence, conclusions (ii) and (iii) of Lemma 4.1 hold. □

C Proof of Theorem 4.4

By definition of Φ,

$$\nabla_{xx}\Phi(x_k, \lambda_k, \mu_k) = \nabla_{xx}\ell(x_k, \bar{\lambda}_k) + \frac{1}{\mu_k} c'(x_k)^* c'(x_k), \tag{C.1}$$

where $\bar{\lambda}_k \stackrel{\text{def}}{=} \bar{\lambda}(x_k, \lambda_k, \mu_k)$. Let s be a nonzero vector such that

$$c'(x_k)s = 0. \tag{C.2}$$

For any such vector, AS4 implies that

$$s^T \nabla_{xx}\Phi(x_k, \lambda_k, \mu_k)s \geq \nu \|s\|^2$$

for some $\nu > 0$, which in turn gives that

$$s^T \nabla_{xx}\ell(x_k, \bar{\lambda}_k)s \geq \nu \|s\|^2,$$

because of (C.1) and (C.2). By continuity of $\nabla_{xx}\ell$ as x_k and $\bar{\lambda}_k$ approach their limits (note that the convergence of $\bar{\lambda}_k$ to λ_* for $k \in \mathcal{K}$ is guaranteed by Theorem 4.3), this ensures that

$$s^T \nabla_{xx}\ell(x_*, \lambda_*)s \geq \nu \|s\|^2$$

for all nonzero s satisfying

$$c'(x_*)s = 0,$$

which implies that x_* is an isolated local solution of (1.1). □

D Proof of Lemma 4.5

We observe that the assumptions of the lemma guarantee that Theorem 4.3 can be used.

Using (2.2) and Taylor's expansion around x_*, we obtain that

$$\begin{aligned}\nabla_x \Phi(x_k, \lambda_k, \mu_k) &= \nabla_x f(x_k) + c'(x_k)^*(\lambda_k + c(x_k)/\mu_k) \\ &= \nabla_x \ell(x_k, \bar{\lambda}_k) \\ &= \nabla_x \ell(x_*, \bar{\lambda}_k) + \nabla_{xx}\ell(x_*, \bar{\lambda}_k)(x_k - x_*) + r_1(x_k, x_*, \bar{\lambda}_k) \\ &= \nabla_x \ell(x_*, \lambda_*) + \nabla_{xx}\ell(x_*, \lambda_*)(x_k - x_*) + c'(x_*)(\bar{\lambda}_k - \lambda_*) \\ &\quad + r_1(x_k, x_*, \bar{\lambda}_k) + r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*),\end{aligned} \tag{D.1}$$

where $\bar{\lambda}_k \stackrel{\text{def}}{=} \bar{\lambda}(x_k, \lambda_k, \mu_k)$,

$$r_1(x_k, x_*, \bar{\lambda}_k) \stackrel{\text{def}}{=} \int_0^1 [\nabla_{xx}\ell(x_* + \sigma(x_k - x_*), \bar{\lambda}_k) - \nabla_{xx}\ell(x_*, \bar{\lambda}_k)](x_k - x_*)\, d\sigma$$

and

$$r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*) \stackrel{\text{def}}{=} (c''(x_*)(x_k - x_*))^*(\bar{\lambda}_k - \lambda_*).$$

The boundedness and Lipschitz continuity of f'' and c'' in a neighbourhood of x_*, together with the convergence of $\bar{\lambda}_k$ to λ_* then imply that

$$\|r_1(x_k, x_*, \bar{\lambda}_k)\| \leq \kappa_7 \|x_k - x_*\|^2, \tag{D.2}$$

and

$$\|r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*)\| \leq \kappa_8 \|x_k - x_*\| \, \|\bar{\lambda}_k - \lambda_*\| \tag{D.3}$$

for some positive constants κ_7 and κ_8. Moreover, using Taylor's expansion again, along with the fact that Theorem 4.3 ensures the equality $c(x_*) = 0$, we obtain that

$$c(x_k) = c'(x_*)(x_k - x_*) + r_3(x_k, x_*), \tag{D.4}$$

where

$$\|r_3(x_k, x_*)\| \leq \kappa_9 \|x_k - x_*\|^2 \tag{D.5}$$

for some positive constant κ_9. Combining (D.1) and (D.4), we obtain

$$\begin{pmatrix} \nabla_{xx}\ell(x_*, \lambda_*) & c'(x_*)^* \\ c'(x_*) & 0 \end{pmatrix} \begin{pmatrix} x_k - x_* \\ \bar{\lambda}_k - \lambda_* \end{pmatrix} = \begin{pmatrix} \nabla_x \Phi(x_k, \lambda_k, \mu_k) - \nabla_x \ell(x_*, \lambda_*) \\ c(x_k) \end{pmatrix} - \begin{pmatrix} r_1 + r_2 \\ r_3 \end{pmatrix}, \tag{D.6}$$

where we have suppressed the arguments of the residuals r_1, r_2 and r_3 for brevity. Note that Theorem 4.3 ensures that $\nabla_x \ell(x_*, \lambda_*) = 0$, such that (D.6) becomes

$$\begin{pmatrix} \nabla_{xx}\ell(x_*, \lambda_*) & c'(x_*)^* \\ c'(x_*) & 0 \end{pmatrix} \begin{pmatrix} x_k - x_* \\ \bar{\lambda}_k - \lambda_* \end{pmatrix} = \begin{pmatrix} \nabla_x \Phi(x_k, \lambda_k, \mu_k) \\ c(x_k) \end{pmatrix} - \begin{pmatrix} r_4 \\ r_3 \end{pmatrix}, \tag{D.7}$$

where $r_4 \stackrel{\text{def}}{=} r_1 + r_2$. Roughly speaking, we now proceed by showing that the right-hand side of this relation is of the order of $\omega_k + \mu_k \|\lambda_k - \lambda_*\|$. We will then ensure that the vector on the left-hand side is of the same size, which is essentially the result we aim to prove. We first deduce from (D.2), (D.3), (D.5) and (4.1) that

$$\left\| \begin{pmatrix} r_4 \\ r_3 \end{pmatrix} \right\| \le \kappa_{10} \|x_k - x_*\|^2 + \kappa_{11} \omega_k \|x_k - x_*\|, \tag{D.8}$$

where $\kappa_{10} = \kappa_7 + \kappa_2 \kappa_8 + \kappa_9$, and $\kappa_{11} = \kappa_1 \kappa_8$. We now bound $c(x_k)$. Using (4.3), we have that

$$\|c(x_k)\| \le \kappa_1 \omega_k \mu_k + \mu_k \|\lambda_k - \lambda_*\| + \kappa_2 \mu_k \|x_k - x_*\|, \tag{D.9}$$

for all k sufficiently large. Hence, by (3.2) and (D.9), we obtain that

$$\left\| \begin{pmatrix} \nabla_x \Phi(x_k, \lambda_k, \mu_k) \\ c(x_k) \end{pmatrix} \right\| \quad \le \quad \omega_k + \kappa_1 \omega_k \mu_k + \mu_k \|\lambda_k - \lambda_*\| + \kappa_2 \mu_k \|x_k - x_*\|. \tag{D.10}$$

By Assumption AS6, the operator on the left-hand side of (D.7) is invertible. Let M be the norm of its inverse. Multiplying both sides of the equation by this inverse and taking norms, we obtain from (D.8) and (D.10) that

$$\begin{aligned} \left\| \begin{pmatrix} x_k - x_* \\ \bar{\lambda}_k - \lambda_* \end{pmatrix} \right\| \quad \le \quad & M[\kappa_{10} \|x_k - x_*\|^2 + \kappa_{11} \omega_k \|x_k - x_*\| + \omega_k + \kappa_1 \omega_k \mu_k \\ & + \mu_k \|\lambda_k - \lambda_*\| + \kappa_2 \mu_k \|x_k - x_*\|]. \end{aligned} \tag{D.11}$$

Suppose now that k is sufficiently large to ensure that

$$\omega_k \le \frac{1}{4M\kappa_{11}} \tag{D.12}$$

and let

$$\bar{\mu} \stackrel{\text{def}}{=} \min \left[\mu_0, \frac{1}{4M\kappa_2} \right]. \tag{D.13}$$

Recall that μ_0 and hence $\bar{\mu} < 1$. Then, if $\mu_k \le \bar{\mu}$, the relations (D.11)–(D.13) give

$$\|x_k - x_*\| \le \frac{1}{2} \|x_k - x_*\| + M[\kappa_{10} \|x_k - x_*\|^2 + \kappa_{12} \omega_k + \mu_k \|\lambda_k - \lambda_*\|], \tag{D.14}$$

where $\kappa_{12} = 1 + \kappa_1$. As $x_k - x_*$, and hence $\|x_k - x_*\|$ converge to zero, we have that

$$\|x_k - x_*\| \le \frac{1}{4M\kappa_{10}} \tag{D.15}$$

for k large enough. Hence inequalities (D.14) and (D.15) yield that

$$\|x_k - x_*\| \le 4M(\kappa_{12} \omega_k + \mu_k \|\lambda_k - \lambda_*\|), \tag{D.16}$$

which is (4.15) where $\kappa_3 \stackrel{\text{def}}{=} 4M\kappa_{12}$ and $\kappa_4 \stackrel{\text{def}}{=} 4M$. Now, using (4.1) and (4.15), we have that

$$\|\bar{\lambda}_k - \lambda_*\| \le \kappa_1 \omega_k + \kappa_2 (\kappa_3 \omega_k + \kappa_4 \mu_k \|\lambda_k - \lambda_*\|),$$

which is (4.16) where $\kappa_5 \stackrel{\text{def}}{=} \kappa_1 + \kappa_2 \kappa_3$ and $\kappa_6 \stackrel{\text{def}}{=} \kappa_2 \kappa_4$. Finally, by (2.1),

$$\|c(x_k)\| = \mu_k \|\bar{\lambda}_k - \lambda_k\| \le \mu_k (\|\bar{\lambda}_k - \lambda_*\| + \|\lambda_k - \lambda_*\|), \tag{D.17}$$

and (4.17) follows from (D.17) and (4.16). □

E Proof of Theorem 4.8

The proof parallels that of Lemma 4.5. First, Theorem 4.7 shows that the penalty parameter μ_k stays bounded away from zero, and thus remains fixed at some value $\mu_{\min} > 0$, for $k \geq k_{\max}$. For all subsequent iterations,

$$\omega_{k+1} = \mu_{\min}\omega_k \text{ and } \eta_{k+1} = \mu_{\min}^{\beta_\eta}\eta_k \tag{E.1}$$

hold. Moreover, we have that $\|c(x_k)\| \leq \gamma_2\eta_k$ for all $k \geq k_{\max}$ sufficiently large. Hence, the bound on the right-hand side of (D.10) may be replaced by $\omega_k + \gamma_2\eta_k$, and thus instead of (D.14)

$$\|x_k - x_*\| \leq M[\kappa_{10}\|x_k - x_*\|^2 + \kappa_{11}\omega_k\|x_k - x_*\| + \omega_k + \gamma_2\eta_k]. \tag{E.2}$$

Therefore, if k is sufficiently large that

$$\omega_k \leq \frac{1}{2M\kappa_{11}} \tag{E.3}$$

and

$$\|x_k - x_*\| \leq \frac{1}{4M\kappa_{10}}, \tag{E.4}$$

inequalities (E.2)–(E.4) can be rearranged to yield

$$\|x_k - x_*\| \leq 4M(\omega_k + \gamma_2\eta_k),$$

that is

$$\|x_k - x_*\| \leq \kappa_{15}\omega_k + \kappa_{16}\eta_k \tag{E.5}$$

where $\kappa_{15} = 4M$ and $\kappa_{16} = 4M\gamma_2$. Since $\beta_\eta < 1$ and $\mu_{\min} < 1$, (E.1) and (E.5) show that x_k converges to x_* at least R-linearly, with R-factor $\mu_{\min}^{\beta_\eta}$. Inequalities (4.1) and (E.5) then guarantee the same property for $\bar{\lambda}(x_k, \lambda_k, \mu_k)$, which in turn guarantees the same property for λ_k, because of (3.6). $\square$